North East Linguistics Society

Volume 16
Issue 1 *NELS 16*

Article 11

1985

A Principle Theory of Categorial Acquisition

W. Neil Elliott
UC - Irvine

Ken Wexler
UC - Irvine

A Principle Theory of Categorial Acquisition

W. Neil Elliott and Ken Wexler

UC - Irvine

0.0. Introduction: "In the beginning was the word ..."

Despite many years of active investigation in the area of language acquisition, the very basic question of how the child learns the grammatical categories of the words in her vocabulary remains unanswered. Most research has begun with the assumption that the learner does indeed learn lexical categories, somehow, and proceeds to attack equally recalcitrant problems. Work in the area of lexical acquisition has concentrated on the development of meanings, e.g., Clark (1973) and Carey (1978) and lexical productivity, e.g., Clark (1982), Bowerman (1982) and Randall (1982), to cite some relevant, if not representative, research. The literature on the acquisition of syntactic categories is sketchy and, we believe, in many cases wrong or misguided.

The paper below has the following structure. First, two dominant theories of the acquisition of syntactic categories are considered in section 1. Section 2 advances the hypothesis that syntactic categories are substantive universals and proposes Feature Realization Principles (FRP's) as a kind of substantive universal connecting grammatical and conceptual representations. Section 3 shows how the FRP's are involved in the acquisition of

nouns. Section 4 argues that verbs are the product of the computational power of devices invested in universal grammar (UG). The result is a theory of categorial acquisition which requires very little extragrammatical information and relies heavily on principles of UG.

1.0. Two Current Theories of Categorial Acquisition.

Any scientific study of a complex process breaks it into individual parts for analysis. For example, it is well-known how linguistic capacities can be analyzed into modules. The theory of language acquisition also breaks down into components. Each component of the process can be analyzed individually. Ultimately we must determine whether the interaction of the components is correct. In many cases, in order to study one important part of the problem we've had to simply assume that another part had been solved, with the appropriate properties emerging. This is natural in science and not to be deplored. However, we have to recognize clearly that a hole has been left, a hole to which we must someday return.

A classic example in the theory of language acquisition has to do with syntactic categorial acquisition. (This does not mean, in our view, that the concepts of particular categories are learned. Rather, it means that the attachment of particular categories to particular formatives is learned.) Coherent theories of the development of a number of relatively complicated structural conditions exist, and these theories, for the most part, assume that the child has already learned the categorial status of a sufficiently large part of his or her vocabulary. This is true of more traditional theories, e.g., the degree-2 theory (Wexler and Culicover (1980)) as it is of more recent theories of parameter setting (e.g., Wexler and Manzini (in press)). The assumption seems reasonable enough--the child at some point has categorial knowledge of a portion of the lexicon. But how is this knowledge attained? In this paper, we return to this gap in the theory.

And gap it is, for there has been, to our knowledge, no successful theory of categorial acquisition. This seems at least mildly disconcerting, because, at first sight, the problem of categorial acquisition seems like a less difficult problem than others that have been more successfully broached.

There have been two major approaches to the theory of category acquisition. Here we can't discuss them in any detail but can only indicate their major features. The first class of approaches is distributional theory. Probably the best-known instantiation of this theory is in Maratsos and Chalkley (1980), but there have been a number of proposals. The idea is that when a child hears words that appear in the same linguistic "context" (e.g., after the or before who) she classes the words together and

then forms categories from classes of words appearing in the same context. In order to show the empirical inadequacies of such an approach, one would have to formalize it and show that it doesn't work. There is no space to do so here, but, in our view, such an analysis will simply not create correct categories--the context in which words of different categories appear are too overlapping. But, without working out the details of any particular distributional theory, there is a fundamental flaw, from our perspective, with all such approaches. Namely, the categories are actually formed as a class of words, so a class of what we call nouns, for example, is formed, and this class can be called anything, say quarks. But the class is just a class of words, and none of the properties of Universal Grammar (UG) relevant to nouns can be brought to bear on this class. For example, there would be no way to apply the Case Filter to phrases headed by elements of this class, because there would be no way to identify these as noun phrases. What distributional theory tells us is that these are words in a class, that is, they appear in similar contexts, but it gives no hint as to why this is just the class that must be marked for case. So distributional theory doesn't arrive at the essential thing that a theory of category formation must arrive at--the appropriate categories for formatives, entering into the intricate structure of UG. In other words, distributional theory does not allow for some of the most essential properties of categories--that they imply particular abstract properties (determined by UG) that are relevant to the words in the categories.

The second major class of existing theories of syntactic category acquisition is the class of semantic bootstrapping theories. These have many precursors, but the more modern, sophisticated form was first articulated by Grimshaw (1981) and it has been extensively developed by Pinker (1984).

The idea of semantic bootstrapping theories is that conceptual or semantic categories are the basis for the induction of grammatical categories, e.g., a person, place or thing is a noun, an event is a verb, a proposition is a sentence.

One immediate problem with such theories is that, except for nouns, they clearly don't work. It is easy to imagine events that are nouns, fire, for example. Semantic bootstrapping theory has to predict a huge number of errors of categorization by the child. The mapping between syntactic and semantic categories simply isn't that close, and we haven't seen a proposal that will overcome these problems.

There is a literature which poses these two approachs against each other. Maratsos (1983) has an extensive and useful critique of the possibility of bootstrapping and Pinker (1984) has a useful critique of distributional theories, although to our knowledge, he doesn't cite the problem of the connection to UG.

Our belief is that none of the existing theories of syntactic category acquisition, whether distributional or bootstrapping, make use of the powerful apparatus of UG that, for all sorts of other reasons concerning language learning, we have to assume that the child has access to. We would like to propose a computational, principled theory of categorial acquisition in which UG plays a powerful role in the creation of category labels for lexical items. In our theory only nouns are bootstrapped, and these in a particular way. After that UG takes over. We believe the output of the process of learning syntactic categories actually connects to the real properties of grammatical classes (as opposed to distributional theories). And we don't have to assume a strong, false connection between syntactic and semantic categories (as opposed to bootstrapping theories). Such a computational, UG-based theory, also makes a lot of sense, a priori. Why shouldn't the learner use UG to learn the attachments of particular formatives to particular categories, given that she has the principles of UG and is using them in sentence construction?

Basically, ours is a formal approach, like distributional theory (and unlike semantic bootstrapping theory), but a formal approach deriving from deep properties of UG, which allows for substantive category labels to be attached to lexical items. Our theory is a computational theory. Bootstrapping theories, which presuppose a direct mapping between conceptual (semantic) and syntactic levels, seem to presuppose a very Gibsonian or "ecological" (noncomputational) view, in which properties are directly "picked up" from the environment. Our theory, on the other hand, is a representational/computational theory in which the acquisition process uses UG strongly as part of a complex learning computation. The principles of UG are used directly. The distributional model is also computational, but relatively non-representational (that is, it has weak, incorrect, representations). We propose that an important insight into the solution to the problem of learning syntactic categories is to be gained from the realization that deep principles of UG must be directly involved. Whatever the empirical status of the particular theory to follow, we believe that principles of UG will play a direct role in the ultimately correct theory of the acquisition of syntactic categories.

2.0. Substantive Universals.

In the scheme of categorial acquisition proposed below, three types of substantive universal are envisioned. Two kinds, syntactic features and lexical redundancy rules, are described in the literature. The third, Feature Realization Principles, bears a resemblance to Grimshaw's (1981) Canonical Structure Realizations,[1] although FRP's are very different in form and intent.

Following Chomsky (1965,1970) and Stowell (1981), one type of substantive universal is a set of syntactic features, combinations of which capture the pretheoretic notion of parts of speech. Syntactic features are divided into two subsets, categorial features and subcategorial features. The inventory of categorial features is given in (1), along with the parts of speech they represent.

1)a	[+N,-V]	noun	b	[-N,+V]	verb
c	[+N,+V]	adjective	d	[-N,-V]	preposition

Subcategorial features combine with categorial features and each other to designate lexical classes. (2) gives some examples.

2)a	[+N,-V,+C(OU)NT]	count noun
	[+N,-V,-CNT]	mass noun
b	[-N,+V,+ACC,θ]	transitive verb
	[-N,+V,θ]	unaccusative (ergative) verb
	[-N,+V,θ_e]	pure intransitive verb

The distinction between categorial features and subcategorial features marks one important difference. Categorial features and combinations of categorial features standing for major classes are conceptually opaque while subcategorial features and subcategories show a tendency towards conceptual transparency. For example, the relation between the class, noun, and any underlying conceptual or ontological category is, at best, obscure. English speakers learn, as part of their language catechism, that a noun stands for a person, place or thing. Ignoring for the sake of argument the apparent heterogeneity of the school definition, the speaker of English knows very well that nouns also stand for actions, events, qualities, locations, numbers and a variety of other ontologically diverse entities. By way of contrast, the grammatical distinction between count and mass nouns is conceptually salient as well,[2] despite controversy over what exactly mass nouns name. For our ontologically naive purposes, it suffices to say that mass nouns refer to stuff and count nouns refer to things. We will return to this fact as an important feature of categorial acquisition.

Beyond universals in the form of syntactic features, we assume two other types. The first of these is familiar, lexical redundancy rules. Lexical redundancy rules serve to relate subcategories to the appropriate category features. The exact form of redundancy rules is beyond the scope of this paper but two plausible candidates are exemplified in (3).

3)a [±C(OU)NT] --> [+N,-V]
b [θ] --> [-N,+V]

(3)a and b are interpreted to mean that if some lexical item is

assigned a marking for the count feature or a θ-role then that lexical item is automatically specified as nominal or verbal, respectively.

The remaining type of substantive universal is at the center of our account of how the language-learner and the developing grammar exploit available conceptual information to determine the syntactic categories of her language. Recall that we observed that the relation between subcategories and conceptual notions is usually (although not unexceptionally) transparent. Suppose then that as part of grammar learning capacity, there is provided a set of functions relating concepts to their grammaticized counterparts, i.e., subcategorial features. (4) exemplifies what we have in mind.

4) G(rigid object) = [+CNT]

(4) asserts that the concept, rigid object, is grammaticized or mapped onto the syntactic feature, [+CNT]. We will call any function like (4) which maps from conceptual structure to a syntactic category a Feature Realization Principle (FRP).

In general, FRP's will have the form in (5), ruling out FRP's of the form in (6).

5) $G(\psi) = \varphi$, where ψ is a feature of conceptual representation and φ is a subcategorial feature of nouns

6)a G(rigid object) = [+N,-V]
b G(agent) = θ_e

(6)a is impossible because categorial features are not included in the range, φ, of FRP's. Similarly, (6)b is excluded because θ_e is not a subcategorial feature associated with nouns. To reiterate the general point, FRP's map conceptual categories onto lower level syntactic features. They do not permit mappings to major class features, [± N, ±V], combinations thereof, or major lexical categories, noun, verb, etc.

(5) is an empirical statement and errs, perhaps, in the direction of being too restrictive. Another alternative is exemplified in (7).

7) G(rigid object) = [+N]

(7) is a special case of an FRP which includes categorial features. Two problems emerge. First, [+N] designates a natural syntactic class comprised of nouns and adjectives. But the conceptual notion, rigid object, does not represent the semantic analogue of the class of substantives. Indeed, it is unlikely that there is any unique semantic correlate. Second, since the class of substantives

includes adjectives, (8) is as likely a candidate for inducing [+N] as (7).

8) G(quality) = [+N]

(8) is, of course, subject to an objection similar to that applied in (7). In general, categorial features are not transparent to the conceptual substrate, an intuition theoretically embodied in the Autonomy Hypothesis.[3]

Revising FRP's to permit combinatorial specifications of classes more closely related to parts of speech as in (9) does not ameliorate the problem.

9)a G(rigid object) = [+N,-V]
b G(event, action) = [-N,+V]
c G(quality) = [+N,+V]
d G(location) = [-N,-V]

A quick examination of the facts shows the proposed FRP's in (9) to be as unreliable as (7) and (8). The lack of correspondences between major form classes and semantic categories remains.

By restricting FRP's to subcategorization features of nouns, several advantages are gained. First, the actual contribution of semantic and other cognitive information to language acquisition is predicted to be small, consistent with the idea that language learning is a biologically supported, task-specific activity. As we will see below, the burden of grammar construction is placed on the computing power of UG itself. Other proposals, such as Pinker (1984), place little if any constraint on bootstrapping mechanisms. The proliferation of bootstrapping hypotheses suggests that, ultimately, a grammar could be constructed piecemeal from other cognitive levels and operations with very little language-specific learning power. Hence, the lack of constraint in such proposals appear to render a specific language acquisition device superfluous. Second, while subcategorization features can be demonstrated to be components of syntactic representation (shown immediately below), they are for the majority of lexical items in a speaker's vocabulary closely associated with their conceptual base. This is as true for the mature speaker as for the developping language learner. Thus, FRP's are simply set of links between grammatical components of cognition and other areas. There is no need to assume that FRP's are a separate subcomponent of the language acquisition device as does Grimshaw (1981) to accommodate the Autonomy Hypothesis. Again, two advantages for FRP's emerge: their extremely restricted scope and their transparent relation to properties of adult grammatical knowledge.

3.0. Acquiring Nouns.

The count/mass distinction in nouns offers an ideal example of how an FRP operates in the acquisition of the features, [+N,-V]. First, we will show that the feature [± CNT] is a subcategorial feature and not simply part of conceptual representation. Then a simple example of nominal acquisition is considered.

3.1. The Syntactic Independence of the Count/Mass Distinction.

The demonstration that [± CNT] is a grammatical feature independent of conceptual representation emerges from attempting to identify the categories, count noun and mass noun, with their respective referents. The classes selected by grammatical and referential criteria correspond over a broad range of nouns, but some divergences are observed. The disparities show that [±CNT] is a feature of nominal morphology independent of noun semantics.

Consider the morphological criteria for distinguishing count ([+CNT]) and mass ([-CNT]) nouns in (10).

10)a Count nouns occur with the plural marker "+s"; mass nouns do not

b Singular count nouns occur with the [-DEF] determiner "a"; mass nouns do not

c Singular count nouns do not occur with the [-DEF] determiner, "sm" or its null allomorph; mass nouns do

d Count nouns occur with numeral modifiers; mass nouns do not

e Count nouns occur with individuating quantifiers (few, a few, many, etc.); mass nouns occur with nonindividuating quantifiers (little, a little, much, etc.)

The criteria in (10) reliably distinguish the two classes of nouns exemplified in (11).

11)	Count nouns:	chair	cat	child
		building	dog	woman
		ball	tree	boy
		auto	bush	aunt
		television	rock	father
	Mass nouns:	mud	barley	milk
		sand	rice	tea
		water	rubber	coffee
		silver	wood	lead
		gold	sludge	styrofoam

The examples considered below in (12) give a partial application of the morphological standard.

12)a	*{Ø,sm}	chair	{Ø,sm}	mud
b	a	chair	*a	mud
c		chairs	*muds	

That is, "chair," a count noun, does not cooccur with "sm" or its null allomorph consonant with (10)c, while it does take the plural marker and "a," satisfying (10)a and b, respectively. "Mud" does not take the plural marker or the [-DEF] specifier "a" while it does appear with "sm" and Ø. As a result, "mud" is classified as a mass noun by distribution. For brevity's sake, we leave it to the reader to confirm the accuracy of the criteria in (10).

An intuitive set of semantic or conceptual criteria for individuals and substances is not difficult to provide; such criteria appear in (13).

13)a A term, t, is individuative iff t counts its extension

b A term, t, is homogenous iff t dissects and cumulates its extension

The terms, "count," "dissect" and "cumulate" receive the definitions in (14).

14)a A term, t, counts its extension, E, iff t is true of every member of E and inapplicable to any arbitrary subdivision of E

b A term, t, dissects its extension, E, iff t is true of every part of E

c A term, t, cumulates its extension, E, iff t is true of any combination of parts of E

The definitions in (14) along with the criteria in (13) discriminate the referents of the examples in (11) into two subclasses, countable things and noncountable stuff. To exemplify, the two nouns "ball" and "rubber" may be predicated of exactly the same set of time/space continua or worms otherwise designated under the description "spongy, red, spherical objects." However, "ball" cannot apply to an arbitrary part or parts of the continua; it must apply to spatial parts which count as balls. "Rubber," on the other hand, is true of any subdivision of the worms down to molecules of rubber and is also true of all their parts jointly, i.e., the stuff of which the balls consist.

Roughly speaking, then, (13)a picks out individuals and

(13)b, substances. Similarly, the count nouns determined under the grammatical criteria in (10) refer to things while the mass nouns refer to stuff. Over the range of examples in (11), the morphological and conceptual criteria converge.

The correlation between the criteria in (10) and (13) is not perfect; conflicts of classification do exist. First, there is a large and notorious list of ambiguous nouns a few of which are exemplified in (15)a.

15)a apple, lamb, chicken, beer, cheese, fruit, ...
b furniture, footwear, luggage, ...
c thing, entity, object, individual, ...

The examples in (15)a are not problematic since they are disambiguated in context, i.e., the presence of a determiner will suffice on any occasion of use to determine whether the noun is assigned an individuative or homogenous reading. We treat these as cases of lexical ambiguity on a par with examples like "bank." Mass terms select a class of determiners and quantifiers distinct from those selected by count nouns. Ambiguous terms are listed twice, that is, once for each type of selectional feature.

(15)b and c provide more interesting cases. The words in (15)b pattern with mass nouns on morphological criteria yet it would be intuitively difficult to identify their referents with homogenous terms. Similarly, while the terms in (15)c are grammatically count, it is not uncommon to find their referents identified as homogenous. For example, "individual" may apply to a ball, a part of a ball, to any arbitrary grouping of parts of a ball and so forth. These divergences serve to show that the feature selecting the morphological distribution of nouns is independent of the semantics of nouns. That is, the Autonomy Hypothesis can be demonstrated at the subcategorial level as well as the categorial level, but only in terms of exceptions to the general rule marked by the FRP.[4] By comparison, such opacity at the level of major class categories is the general case rather than the exception.

3.2. The Operation of a Feature Realization Principle.

The relevance of the dual nature of the count/mass distinction emerges when it is embedded in a theory of language acquisition. Suppose, following Spelke (1982), the child has sufficient cognitive sophistication to distingish her world into objects and nonobjects. Among nonobjects will be substances, qualities, actions and events. If something is taken by the child to be an object, then in conceptual structure it will be assigned an individuative term marked by [+IND]. The property of individuation is linked to the corresponding grammatical category by the FRP in (16), a replacement for the less formal (5).

16) G([+IND]) = [+CNT]

[+CNT] is, in turn, associated with the feature complex, [+N,-V], by the redundancy rule in (3)a.

Consider a concrete example. The child hears the utterance, "That's a ball," in the presence of a red, round object. She associates the constituent "ball" with the individuating conceptual term in (17)a.

17)a ROUND( )[5]
[+IND]
b [/bɔl/, ROUND( )]
[+IND]

(17)b represents primary linguistic data which will be utilized in the construction of the lexicon. The FRP will assign the datum in (17)b the subcategorial feature, [+CNT]. The lexical item in (18) will result from the interaction of the redundancy rule in (3)a with the FRP in (16).

18) [/bɔl/, [+N,-V,+CNT,...], ROUND( )]
[+IND]

The word "ball" is now appropriately categorized as a count noun. Either simultaneous or renewed experience with the term can be expected to fill in other features and add further information to the conceptual representation of recurring experiences with small spheres.

4.0. Computing the Features, [+N,-V].

Given an utterance like (19), the child's knowledge that "dog" is an active object will interact with UG to permit the inference that "run" is a verb.

19) The dog ran

The process will be roughly the following. (20) informally represents the FRP for animacy.

20) G(active object) = [+ANI(MATE)]

The FRP permits the learner to associate her conceptual knowledge of dogs with a subcategorial feature that occurs as a nominal subcategory via the redundancy rule in (21)a.

21)a [+ANI] --> [+N,-V]
b [/dɔg/, [+N,-V,+ANI,...], ACTIVE OBJECT]

Thus, the word "dog" will receive the lexical encoding in (21)b. So

far, the process has required nothing other than available conceptual information and substantive universals of grammar.

Wanner and Gleitman (1982) argue that the child uses her knowledge of aspects of prosodic structure to assign constituent boundaries to the utterance in (19) as shown in (22).

22) [[The #dog#][#ran#]]

Stressed syllables form a phonetic nucleus to which word boundaries, #...#, can be assigned and intonational contour provides a reliable clue to phrasal constituents, [...]. The prosodically identified constituents are unlabelled for syntactic category.

The minimal constituent structure assigned under prosody is both elaborated and constrained by X-bar Theory. The further assignment of structure to (22) is limited to the representation in (23), assuming that phrasal projections are maximal and that the FRP in (20) eventuates in the partial labelling of categories shown.

23) $[_{Z"}[_{N"}$ The #dog#] Z $[_{Y"}$ #ran#]]

Because X-bar Theory imposes the requirement that every maximal projection have a lexical head, the child is forced to assume some abstract structure, viz., Z, associated with utterances. However, the learner is not without some ability to infer the syntactic properties of the abstract element. At this point, the computational problem reduces to using principles of UG to infer the lexical or inflectional (in Borer's (1983) sense) properties of Z and the head of Y". Case Theory and θ-Theory play important roles.

The Case Filter in (24) along with the notion of Case assignment formalized in (25) permits the learner to infer that Z must have Case marking properties.

24) *NP if NP is lexicalized and is not Casemarked

25) α assigns Case to β iff
i) α is a Case assigner
and ii) α governs β

Since case assigners are typically lexical constituents and Z is the only nonphrasal constituent governing NP, it follows that Z is a Case assigner and one syntactic property of its lexical listing is that it will exhibit a Case feature as shown in (26).

26) [/-∅/, [Z,+Case,...], ...]

Z will turn out to be INFL in the adult lexicon but the means by

which the learner arrives at this conclusion is beyond the scope of this paper.

θ-Theory will play a somewhat more complex role. Since the child knows that "the dog" is an NP, she can infer that it must have a θ-role as required by the θ-Criterion given in (27).

27) Every NP must have a θ-role and every θ-role must be assigned to some NP

The principles of θ-role assignment in (28) would lead the child to a potentially incorrect conclusion about the thematic properties of Z without some further constraint.

28) α assigns a θ-role to β iff
i) α has a θ-role to assign
and ii) α governs β

Since Z governs the NP in (23), the learner could conclude that Z does have a θ-role to assign. Whatever role Z, as the constituent INFL in the adult grammar, plays in θ-role assignment, it is not assumed to possess a independent θ-role and the child's inference would be erroneous.

The source of error can be avoided if a generalization of the Thematic Inference Principle assumed in Wexler (1982) and Borer and Wexler (in press) is adopted. The generalized principle is stated in (29).

29) A child inferring θ-roles from a situation will assume a θ-role iff there are overt constituents which are construed as being in a thematic relation

The narrower principle proposed originally by Wexler (1982) requires a close relation between linguistic structure and conceptual representation by insuring that the child does not infer abstract grammatical properties such as implicit θ-roles from information available in the situation but unexpressed in the utterance. The expanded principle in (29) extends the guarantee to avoid the postulation of abstract role assigners on the basis of relations in the situation that are not represented by elements of the sentence.

The principle in (29) applies to the situation where the language learner is attempting to infer an θ-role for "the dog." That is, the θ-role assigned to the NP will be inferred from the situation only if there are two overt constituents which can be construed as having a thematic relation. "The dog" and "ran" fulfill the requirement but not "the dog" and Z. Z is eliminated as a possible θ-role assigner. Once Z is ruled out it is not difficult to see that the child will assume "ran" as the constituent

assigning the θ-role.

On our assumptions so far the learner is able to associate a θ-role with the constituent [$_{Y"}$ ran]. However, this leaves the lexical categorization of "ran" open in two ways. First, the verb has not been assigned any categorial features. Second, the θ-role assigned by "ran" could play rather different roles in subcategorization depending on whether it is internal or external. The first deficiency is trivially remedied by the redundancy feature posited in (3)b linking θ-roles to the feature complex, [-N,+V]. Resolving the second requires some further assumptions.

Although the child is restricted to inferring only one θ-role from the situation by the Thematic Inference Principle in (29), she cannot determine from that inference alone whether the verb is transitive, unaccusative or intransitive. Several ancillary assumptions are required.

One crucial assumption, the Maturational Hypothesis, stated in (30), originates in Borer and Wexler (in press) where it is given impressive empirical and theoretical support.

30) Early grammars do not permit the nonlocal assignment of subcategorial features

Given (30), one possibility is ruled out immediately. Suppose the child arbitrarily assigned "run" an internal θ-role. As a consequence, the underlying representation of (23) would, at some level of grammatical structure, have to be (31), since only external θ-roles can be locally assigned to subjects (by VP).

31) [$_{Z"}$ e Z [$_{V"}$ ran [$_{N"}$ the #dog#]]]

But if (31) underlies (23), then (32) must be the S-structure associated with (23).

32) [$_{Z"}$ [$_{N"}$ The dog]$_i$ Z [$_{V"}$ ran e_i]]

However, the process of A-chain formation under movement presupposed by the structure in (32) is enjoined under the Maturational Hypothesis because it requires the nonlocal assignment of subcategorial features. Hence, it follows that the child cannot erroneously categorize "run" as an unaccusative verb. Further support for this hypothesis comes from Borer and Wexler's (in press) argument that young children actually miscategorize unaccusatives as pure intransitives.

The Maturational Hypothesis interacts with other principles of UG in such a way that the language learner can deduce that "run" is not a transitive verb. The further principles of UG that must be brought into play are expressed in Burzio's Generalization, (33).

33) If a verb assigns [+ACC] Case, then it also assigns an external θ-role

Suppose that the language learner arbitrarily assumes the verb in question assigns [+ACC]. Then, for the case of "run," she will correctly infer that it assigns an external θ-role. Given that "the dog" is the only plausible candidate for being assigned a θ-role in (23), the Maturational Hypothesis forces a dilemma. [+ACC] can only be assigned to an NP locally, i.e., to an NP in direct object position. The external θ-role can only be locally assigned to subject position. There are two possibilities. Either the NP is assigned Case in direct object position and moved to subject position to receive the θ-role or it is generated in subject position, receiving the external θ-role from VP and Case from Z. The former alternative is ruled out by the Maturational Hypothesis. The latter option results in the representation in (34).

34) $[_{Z''}\ [_{N''}\ \text{The dog}]\ Z\ [_{V''}\ \text{ran}\ [_{N''}\ e]]]$
$[_{N''}\ [+\text{NOM}]\ [\theta_e\ ]]$ $[_{V''}\ [+\text{ACC}]]$

The Projection Principle has the effect of requiring an NP complement to which the subcategorial feature, [+ACC], can be assigned. Presumably, the complement is construed as an argument. But if this is so, then the θ-Criterion will be violated. That is, the Thematic Inference Principle prevents the learner from inferring more than one θ-role since there is only one overt argument, the subject. But if the learner infers only one θ-role and it is assigned to the subject, then it follows that the NP complement violates the θ-Criterion by failing to have a θ-role. Since both possibilities are ruled out by a combination of UG, the Maturational Hypothesis and the Thematic Inference Principle, the child must conclude that "run" does not assign accusative case and is not a transitive verb.

In effect, principles of UG, the Thematic Inference Principle, and the Maturational Hypothesis interact to permit the child to compute correctly the grammatical properties of a pure intransitive verb. In the service of brevity, we will not discuss the acquisition of other classes of verbs. We assume that the process can be straightforwardly extrapolated from the example given.

5.0. Conclusion.

To recapitulate briefly, we have argued that distributional and semantic theories of the acquisition of grammatical categories are inadequate. The former fails to solve the problem of induction, while the latter is wildly unrestricted. Our own proposal assumes that grammatical categories are given as a set of substantive

universals in UG. There is a limited amount of mapping between conceptual representation and subcategories of nouns expressed by FRP's. The features associated with verbs are computed by the interaction of the lexicon and principles of UG. Overall, the system that results requires very little conceptual information to begin the task of constructing a set of lexical categories for the language being learned.

FOOTNOTES

[1]Although space prevents us from discussing this claim and its merits in any detail, two points are important enough to mention. First, FRP's are far more restricted in form than Canonical Structure Realizations (CSR's). Second, FRP's are conceived of as being part of the grammar while the CSR's in Grimshaw's (1981) proposal are an independent subcomponent of the language acquisition device.

[2]Cf. below, section 3., (13), for a coherent semantic definition of the notions mass and count.

[3]It should be stressed that we do not find the alternative expressed in (7) and (8) unappealing. For example, the child could learn that "ball" is [+N] via (7) and [-V] because it fails to assign a θ-role. "Red," on the other hand, would be characterized by (8) as [+N] because it is a quality and as [+V] because it can occur as a θ-role assigning predicate. Kean (p.c.) points out that the undesireable consequences of (7) and (8) disappear for some languages, for example, Walpiri, where adjectives and nouns seem to be distinguished only by function. However, another kind of problem emerges. That is, Dixon (1977) observes languages which lack a productive class of adjectives but those languages appear to collapse the categories with verbs rather than nouns. Obviously, on this option, much remains to be worked out.

[4]These exceptions are the product of the overgeneralization of the morphological component and make the prediction that such cases will not be learned until the morphology distinguishing mass and count nouns is fully productive.

[5]Although we wish to avoid taking a position on the formal nature of conceptual representation, ROUND( ) can be taken as a monadic predicate true of the individuals in its extension. Other concepts are expressed by n-adic predicates satisfied by the n-tuples in their extensions. Such predicates may be complex, i.e., composed of other predicates. Another alternative would be to conceive of conceptual representation in terms of semantic features, say, following Katz (1972).

REFERENCES

Bowerman, M. (1982). "Reorganizational Processes in Lexical and Syntactic Development," in Language Acquisition: The State of the Art, ed. E. Wanner and L. Gleitman. Cambridge: Cambridge University Press.

Borer, H. (1983). Parametric Syntax: Case Studies in Semitic and Romance Languages. Dordrecht: Foris Publications.

Borer, H. and K. Wexler (in press). "The Maturation of Syntax," in Parameters and Linguistic Theory, ed. T. Roeper and E. Williams.

Carey, S. (1978). "The Child as Word Learner," in Linguistic Theory and Psychological Reality, ed. M. Halle, J. Bresnan, and G. Miller . Cambridge: MIT Press.

Clark, E. (1973). "What's in a Word? On the Child's Acquisition of Semantics in His First Language," in Cognitive Development and the Acquisition of Language, ed. T.E. Moore. New York: Academic Press.

Clark, E. (1982). "The Young Word-Maker: A Case Study of Innovation in the Child's Lexicon," in Language Acquisition: The State of the Art, ed. E. Wanner and L. Gleitman. Cambridge: Cambridge University Press.

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge: MIT Press.

Chomsky, N. (1970). "Remarks on Nominalization," in Studies on Semantics in Generative Grammar. The Hague: Mouton and Company.

Dixon, R. (1977). "Where Have All the Adjectives Gone?," Studies in Language 1:19-80.

Grimshaw, J. (1981). "Form, Function and the Language Acquisition Device," in The Logical Problem of Language Acquisition," ed. C.L. Baker and J.J. McCarthy. Cambridge: MIT Press.

Jackendoff, R. (1977). X-bar Syntax. Cambridge: MIT Press.

Katz, J. (1972). Semantic Theory. New York: Harper and Row, Publishers.

Maratsos, M.P. (1983). "Some Current Issues in the Study of the Acquisition of Grammar," in The Handbook of Child Development.

Maratsos, M.P. and M.A. Chalkley (1980). "The Internal Language of Children's Syntax: The Ontogenesis and Representation of Syntactic Categories," in Children's Language, ed. K. Nelson. New York: Gardner Press.

Pinker, S. (1984). Language Learnability and Language Development. Cambridge: Harvard University Press.

Randall, J. (1982). Morphological Structure and Language Acquisition. Ph.D. dissertation, University of Massachusetts, Amherst.

Spelke, E. (1982). "Perceptual Knowledge of Objects in Infancy," in On Mental Representations, ed. J. Mehler, M. Garrett, and E. Walker. Hillsdale,N.J.: Erlbaum.

Stowell, T. (1981). The Origins of Phrase Structure. Unpublished MIT doctoral dissertation.

Wanner, E. and L. Gleitman, ed. (1982). "Introduction," in Language Acquisition: The State of the Art. Cambridge: Cambridge University Press.

Wexler, K. (1982). "A Principle Theory for Language Acquisition," in Language Acquisition: The State of the Art, ed. E. Wanner and L. Gleitman. Cambridge: Cambridge University Press.

Wexler, K. and P. Culicover (1980). Formal Principles of Language Acquisition. Cambridge: MIT Press.

Wexler, K. and R. Manzini (in press). "Parameters and Learnability in Binding Theory," in Parameters and Linguistic Theory, ed. T. Roeper and E. Williams.